FG

12

AD A048567

001813-5-F

UNIVERSITY OF SOUTHERN CALIFORNIA

USC

social science research institute

TECHNICAL REPORT

RESEARCH ON THE TECHNOLOGY OF INFERENCE AND DECISION

WARD EDWARDS
DAVID A. SEAVER

SPONSORED BY:
ADVANCED RESEARCH PROJECTS AGENCY
DEPARTMENT OF DEFENSE

MONITORED BY:
ENGINEERING PSYCHOLOGY PROGRAMS
OFFICE OF NAVAL RESEARCH
CONTRACT NO. N00014-76-C-0074, ARPA

OCTOBER 1976

SSRI RESEARCH REPORT 76-7

Social Science Research Institute
University of Southern California
Los Angeles, California 90007
213-741-6955

The Social Science Research Institute of the University of Southern California was founded on July 1, 1972 to permit USC scientists to bring their scientific and technological skills to bear on social and public policy problems. Its staff members include faculty and graduate students from many of the Departments and Schools of the University.

SSRI's research activities, supported in part from University funds and in part by various sponsors, range from extremely basic to relatively applied. Most SSRI projects mix both kinds of goals — that is, they contribute to fundamental knowledge in the field of a social problem, and in doing so, help to cope with that problem. Typically, SSRI programs are interdisciplinary, drawing not only on its own staff but on the talents of others within the USC community. Each continuing program is composed of several projects; these change from time to time depending on staff and sponsor interest.

At present, SSRI has six programs:

Program for research on crime control. Typical projects include evaluation of a federal program for decriminalization of juvenile status offenders; and development of an inventory of the contents and quality of the information held by criminal justice agencies in Los Angeles County.

Program for the study of dispute resolution policy. Typical projects include collection and analysis of national statistical data concerning the size, cost, and performance of present dispute resolution systems in six other countries; and detailed study of some 30 alternatives to present U.S. criminal justice procedures.

Program for research on desegregation. The present goal of this program is to study the effects of language, physical attractiveness, and community contact on acceptance of minority children in white schools and on their scholastic performance.

Program for research on decision analysis. Typical projects include study of elicitation methods for continous probability distributions; and development of a multi-attribute utility measurement method for evaluating social programs.

Program for research on rights of the mentally ill. This program is studying procedures used in Los Angeles Courts to determine whether a non-criminal mentally ill person is sufficiently dangerous to others or to himself to justify his involuntary custodial confinement.

Program for data research. Typical projects include development of techniques for estimating small-area population sizes between censuses; and development of crime indicators for use in criminal justice system planning.

SSRI anticipates that new programs will be added and old ones will be redefined from time to time. For further information, publications, and the like, write or phone the Director, Professor Ward Edwards, at the address given above.

9 USC-001813-5-7

Final Technical Report, SSRI-76-7

Jul 75 - Sep 76, RR- cover

6

Research on the Technology of Inference and Decision,

10

Ward Edwards ~~and~~ David A. Seaver

Social Science Research Institute
University of Southern California

11 October 1976

12 37p.

Sponsored by

15

Defense Advanced Research Projects Agency
ARPA Order ~~No.~~ 3052

N00014-76-C-0074,

Decisions + Designs Inc. McLean, Va

~~392 066~~

390 664

mt

Summary

This report summarizes 15 months of research on the technology of inference and decision, including topics such as measurement and validation of multiattribute utilities, equal weights in multiattribute utility models, group processes for probability and utility assessment, assessing very small probabilities, and the effects of response scales on judgments of uncertainty. Research on these, and other, topics is reported in eight technical reports, summaries of which are included at the end of this report. Here, we explain how research on these specific topics integrates into an overall program of research.

The major themes of our research are guided by difficulties encountered in the application of decision analytic techniques to real-world decision problems. Probably the most important theme is the validation of multiattribute utility models and assessment procedures. All multiattribute utilities depend on two factors: the specific model used and the expert judgments provided as inputs to the model. We believe a tradeoff between errors in these two factors determines the validity of the resulting utilities. Simpler models that are more likely to misrepresent "true" preferences require fewer and simpler judgmental inputs, thereby reducing the judgmental error.

We have used both experimentation and simulation to find where these errors exist, how extensive they are, and how they affect the final evaluative process. An experimental study compared preferences expressed by subjects with those predicted by various multiattribute utility models and found the preferences consistently violated the additive model. But the additive model cannot be rejected because its use may reduce the judgmental error in the utilities.

Simulation has also proved to be a useful tool for investigating the validity problem. It allows us to define "truth" and compare various approximations to the defined truth. To date, we have focussed primarily on comparing additive models with equal and differential weights. Models with equal weights approximate differentially weighted models quite well when the inter-attribute correlations of choice alternatives are all positive. However, when many of these correlations are negative, the typical case for multiattribute choice problems, the approximations are no longer very good.

The second theme of our research, the assessment of group probabilities and utilities, developed from problems encountered when groups rather than single individuals function as decision makers. Theoretical results indicate no completely satisfactory method exists for combining the individual probabilities and utilities of group members into group probabilities and utilities. We have reviewed the advantages and disadvantages of both mathematical and behavioral methods that have been suggested for forming these group judgments. In the probability domain, simple averaging seems to be the most practical mathematical approach. Behavioral approaches depending on structured communication and/or interaction among group members also appear to be useful. The use of multiattribute utility measurement may reduce disagreement about utilities. One particular form based on a simple multiattribute rating technique, SMART, has been used successfully in several contexts where "public" values were needed for decision making.

The final theme of our research is the elicitation and quantification of uncertainty. Much of this research is oriented toward finding practical solutions to problems exposed by previous research. We conducted three experiments examining biases in different situations requiring judgments of uncertainty and how different elicitation procedures could reduce those biases. One experiment investigating the effects of response scales on odds judgments indicated that responses on logarithmically spaced scales were more veridical than responses on linearly spaced scales or simple written responses with no scale. A second study confirmed the existence of several biases in assessed subjective probability distributions. The extent of these biases was dependent to some degree on the procedure used to elicit the distributions. Another study suggested that a new type of judgment might be used to reduce conservatism in probabilistic information processing. Posterior odds calculated via Bayes' Theorem from subjects' judgments of "average certainty" were very nearly veridical.

The assessment of very small probabilities has become a primary concern under this theme. We have identified a procedure using "marker" events that we believe may be a viable alternative to fault trees and direct judgments currently in use. An extensive program of experimental research may be necessary to develop this method.

Contents

Summary i

Acknowledgments iv

Disclaimer v

I. Introduction 1

II. A Technical Overview 2

A. Validation of Multiattribute Models and Assessment Procedures 2

1. Behavioral tests 3

2. Simulation and approximation 4

B. Assessment of Group Probabilities and Group Utilities . . 7

1. Group probabilities 8

2. Group utilities 9

C. Elicitation and Quantification of Uncertainty 10

1. The effect of response scales on odds judgments . . . 11

2. Biases in subjective probability distributions . . . 11

3. Averaging as a means of probabilistic inference . . . 12

4. Assessing very small probabilities 13

III. References 15

IV. Summaries of Technical Reports 17

Distribution List 30

DD Form 1473 . . . .

Acknowledgments

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was Monitored by the Office of Naval Research under Contract N00014-76-C-0074 under subcontract from Decisions and Designs, Inc.

The authors would like to thank Lee Eils, Tsuneko Fujii, Peter Gardiner, J. Robert Newman, William Stillwell, and Detlof von Winterfeldt who made many valuable contributions to the research summarized in this report.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

I. Introduction

This Final Report summarizes the work by the Social Science Research Institute, University of Southern California, on subcontract P. O. 75-030-0711 from Decisions and Designs, Inc., prime contract N00014-76-C-0074 from the Advanced Research Projects Agency, monitored by the Engineering Psychology Programs, Office of Naval Research. The research conducted during this contract period from July 1, 1975 to September 30, 1976 under the direction of Professor Ward Edwards, the Principal Investigator, was part of an ongoing program of Research on the Technology of Inference and Decision. Edwards (1973, 1975) summarizes previous research.

The proposal leading to this subcontract called for research on five specific topics: measurement and validation of multiattribute utilities, equal weights in multiattribute utility models, group processes for probability and utility assessment, assessing very small probabilities, and the effects of response scales on judgments of uncertainty. Our research on these, and other, topics is reported in eight technical reports which have been produced or are now being prepared. Summaries of these technical reports appear at the end of this report.

The purpose of this report is to explain how this research integrates into an overall program of research on decision technology. Thus, we do not report in detail findings that are set forth in the self-contained technical reports. Only major findings are reviewed along with ongoing research and future research possibilities suggested by our current research.

II. A Technical Overview

The major themes of our research are guided by difficulties encountered in the application of decision analytic techniques to real-world decision problems. Although all the research is applications-oriented, both theoretical and pragmatic questions are addressed. The first, and we feel most important, theme is the validation of multiattribute utility models and assessment procedures. We have considered several quite divergent, although related, approaches to this topic ranging from measuring goodness of approximation via simulation to testing the behavioral assumptions underlying various models.

The second topic of research comes as a response to the question of how groups should make decisions. The practical importance of this question is obvious, yet no entirely satisfactory answer is known. We feel sure that groups, like individuals, should make decisions on the basis of probabilities and utilities. But how are group probabilities and utilities to be determined? We have reviewed the current state-of-the-art looking for useful guidelines.

Our final topic of research, the elicitation and quantification of uncertainty, has been studied longer than the other topics, and so, though equally important, seems more familiar and perhaps less puzzling. One particular area of this research that is new and exciting is the assessment of very small probabilities. We have explored how judgments expressing uncertainty can be obtained in a variety of uncertain situations, and the effects of the elicitation procedures on these judgments.

II.A. Validation of Multiattribute Utility Models and Assessment Procedures

As the development and application of multiattribute utility models has become prominent in decision analysis, the problem of how to validate these models has become very pressing. Do the utilities that are the outputs of such models actually represent the preferences of the decision maker who has been modeled? Since there are no known "true" utilities against which to compare the utilities of the model, no simple solution to this problem exists.

Multiattribute utilities depend on two factors: the specific model used and the expert judgments provided as inputs to the model. Each of these factors contributes to the validity of the resulting utilities. Either or both may be erroneous to some degree. We need to know when both modeling error and judgmental error occur, how severe the error is, and ultimately, whether or not the error makes a difference in the decision making process.

II.A.1. Behavioral tests. Our research suggests partial answers to some of these questions. For example, one way of determining whether a given model is appropriate or not is to test the behavioral implications of the model. Different models imply different preference patterns. Comparing model predictions with actual preference patterns can show which models the preference patterns violate. Note, however, that this is a one-sided test: models can only be shown to be wrong, but never proven to be correct.

von Winterfeldt (see summary No. 5), using this approach, compared the appropriateness of several models in a risky multiattribute choice situation. His most notable finding was a consistent violation of the independence assumption that is necessary for the preferences to be consistent with the additive model. Subjects consistently expressed what has been termed multivariate risk aversion; that is, a preference for gambles with outcomes more or less balanced across attributes. For example, with two-attribute outcomes subjects preferred gambles in which they could win something in both attributes to gambles in which they could win a lot in one attribute and nothing in the other.

These results would seem to indicate that the additive model is inappropriate in choice situations such as those studied by von Winterfeldt. However, before rejecting the additive model, several questions must be answered. First, the expressed preferences of the subjects that violated the additive model must be shown not to be the result of judgmental error. As von Winterfeldt points out, two types of errors may make the data uninterpretable or meaningless: inconsistency in preferences or the use of simplifying strategies that are consistent but do not actually represent the preferences of the subject. von Winterfeldt indicates that neither of these problems are apparent in his data.

A second question is how generalizable are the results. If the preference patterns expressed by subjects in this experiment are not similar to preference patterns in a wide variety of choice situations, the lack of generalizability would restrict the usefulness of the findings. However, the subjects in this study did exhibit preference patterns that are characteristic of many choice situations: namely, marginally decreasing single attribute riskless utility, nonlinear tradeoffs between attributes, and risk aversion. Therefore, similar preference patterns can be expected in situations where these characteristics are evident suggesting that the additive model is not appropriate in a wide variety of contexts.

A final consideration must be the consequences of using the additive model when it is inappropriate. At first, this seems a rather unusual question. Assuming there is some other model that has not been shown to be inappropriate, why even consider using an obviously inappropriate model? The problem however, is not that simple. The reason for considering use of an inappropriate model is that it may reduce the judgmental error entering into the resulting utilities. For example, the additive model requires fewer judgmental parameters than most other multiattribute models. Assuming there is some judgmental error in all preference judgments, reducing the number of judgmental parameters should reduce the total judgmental error. The question, then, is whether or not the reduced judgmental error compensates for the modeling error. As models become less likely to be inappropriate, they become more complex both mathematically and in the judgmental inputs necessary. Thus, the tradeoff between model appropriateness and judgmental error seems likely to range over the entire range of possible models. Just what this tradeoff is, and, therefore, which model should be used in a given situation remains to be worked out.

II.A.2. Simulation and approximation. The use of simulation to show the goodness of approximations is another approach we have taken to investigate whether modeling and judgmental errors make a difference in decision making. Newman, Seaver, and Edwards (see summary No. 3) have developed a general purpose simulation technique that can be used to explore a wide variety of questions. Many of our ideas grew out of some analytic work by Einhorn and Hogarth (1975) and Wainer (1976) showing that under certain very broad conditions, linear models with equal weights would predict as well or nearly as well as linear models with weights derived through multiple regression. These findings aroused our curiosity, since the form of multiattribute utility measurement advocated by Edwards and his colleagues (see Edwards, 1972) includes use of an additive model with the same mathematical form as the multiple regression model: the sum of single attribute utilities weighted by the importance of the attribute. The obvious question is: if equal weights work in a multiple regression model, wouldn't they also work in a multiattribute utility model. Should the answer to this question be yes, judgments of the importance of each attribute would no longer be necessary, eliminating one possible source of judgmental error. Again the question is one of trading off modeling error with judgmental error.

The first simulation study by Newman et al. served as a test of the general simulation process. In it, we compared the predictive ability of equal weight models with differential weight models in which the weights were derived by multiple regression. As expected, the differential weight models outperformed the equal weight models, although only slightly.

The difference in predictive ability decreased as sample sizes decreased and as measurement error was added to the independent and dependent variables. We must note, however, that the conditions of this simulation were the most favorable possible for the regression weights: normally distributed predictor variables, relatively large sample sizes, moderately high positive intercorrelations among predictor variables, and a small number of predictor variables (3).

Although the formal models are the same for multiple regression and additive multiattribute utilities, there are certain differences. For example, no criterion exists against which to compare the output of the multiattribute utility model. In addition, most of the above mentioned variables that characterize situations where additive models with weights estimated via multiple regression will work well are not usually found in choice situations where multiattribute utility models are used. The distribution of alternatives on each attribute will not typically be normal. Often the number of alternatives (sample size) will be small. The intercorrelations among the attributes will necessarily be negative on the average, if only admissible alternatives are considered.

Thus, in order to get a better understanding of how equal weights might work in multiattribute utility models, we conducted a second *simulation study*. This study compared three two-attributed models: an equally weighted additive model, a differentially weighted additive model, and a differentially weighted model that included a cross-product term in addition to the additive term of each attribute. Consequently, we could look at not only approximating differential weights with equal weights, but also at approximating a nonadditive model by an additive one with either equal or differential weights. In addition, the effects of both positive and negative correlations between attributes were examined. The results with a positive intercorrelation, as expected, showed the equal weight additive model to be a very good approximation to the differential weight additive model. Both additive models were quite good approximations to the nonadditive model when the cross-product term had a relatively low weight compared with the additive terms. However, as the relative weight given the cross-product term increased, both additive approximations worsened with neither seeming to be better than the other.

The results with a negative intercorrelation were strikingly different. The approximations were all considerably worse than they had been with a positive intercorrelations. As the differential weights in the additive model became more dissimilar (2:1 ratio of larger to smaller) the equal weight approximation became very poor. Moreover, neither additive model approximated the nonadditive model well with the equally

weighted model being slightly worse. This simulation pointed to the need for further study of the use of equal weights in multiattribute utility models. All previous investigations, both analytic and simulation, considered only the case of positive intercorrelations among attributes. Our study showed that negative intercorrelations greatly change the results.

Newman (see summary No. 4) followed up the simulation studies by comparing weighting schemes in a realistic choice setting. The Automobile Club of Southern California has a Target Car program which uses a procedure similar to multiattribute utility measurement to rate how well automobiles meet an optimal design. Newman compared the Automobile Club rankings with a model that gave equal weights to the eleven attributes defined as important by Auto Club engineers and members. Naturally, there were negative correlations among many of the attributes; for example, between large interior size and small exterior size. The results of this study showed that the equal weight model would lead to somewhat different rankings of the automobiles considered. Although, the rankings were not too dissimilar (rank order correlation = .77), differences were substantial enough to suggest use of the equal weight model might bring about the choice of an automobile with less utility than the one ranked best by the Auto Club. In particular, the automobile ranked first by the equal weight model tied for eighth in the Auto Club rankings.

Our studies of approximations have just scratched the surface of the work that needs to be done on this topic. For example, most of the previous work, ours included, has depended on correlations as the measure of goodness of approximation. Although this is a good measure of overall similarity, it may not answer the real question with which we are concerned: what is the loss in utility? New measures need to be tried and simulation will allow us to try them. Using simulation we can define "truth" and show just how much utility would be lost by using various approximations in different choice situations.

In addition, we need a much more thorough analysis of the effects of inter-attribute correlations, and a comparison of additive models with a much wider range of nonadditive models. A topic that we have not yet explored is the use of linear approximations to nonlinear single attribute utility functions. Fischer (1972) reports some results that suggest such approximations may be quite good. We hope to investigate this further, since, again, elimination of the judgments necessary to derive the shape of the single attribute utility curves would reduce the possibility of judgmental error.

As we began our investigations of approximations, we envisioned the simplest of all possible multiattribute utility models as being one in which people had to judge only whether an attribute contributed positively, negatively, or not at all to the overall utility and the best and worst feasible outcomes on each attribute. These judgments could then be used to determine an overall utility for each alternative using linear single attribute utility functions, an additive model, and equal weights. Our work has suggested that such an approximation may not really work in a variety of situations, but we still believe that some degree of approximation is useful in most multiattribute models, and will continue working to show what types of approximations will work in which situations.

Although simulation is a very useful tool for studying approximations, it obviously cannot cover every possible situation. Therefore, a formal analytic theory would be very useful, not only to provide guidelines for using approximations, but also to provide some unification to this diversified collection of ideas. To date, very little has been done analytically that is of practical use. Fishburn (1976) has made an important first step in showing how formal mathematical approximation theory can be applied. We plan to continue efforts along these lines.

II.B. Assessment of Group Probabilities and Group Utilities

Groups rather than individuals often serve as decision makers. Believing as we do that decisions should be made to maximize expected utility, this means that the probabilities and utilities of the decision making group must be determined. A single set of probabilities and utilities that somehow represent the thinking of the group members is needed. Certainly, all members of the group will not agree, so the problem becomes how to aggregate the diverse opinions into a single "group opinion" that represents all group members, and, moreover, is acceptable to the group as a basis for making decisions.

The mathematical and social psychological difficulties of this process are well known. Mathematical problems arise when the obvious approach of using some mathematical combination rule (for example, averaging) is tried. Arrow (1951) has proved a theorem stating that there is no such rule for combining individual preferences (utilities) into a group preference that satisfies a set of reasonable and desirable conditions. A similar theorem was proved in the probability domain by Dalkey (1972).

Social psychological difficulties produce problems when behavioral methods are used to arrive at group probabilities and utilities (Collins and Guetzkow, 1964). Dominance by certain group members because of personality or status may control the group output. Or the group may worry more about achieving an agreement than about the quality of the judgment.

Seaver (see summary No. 2) has reviewed the advantages and disadvantages of several possible methods for forming group probabilities and utilities. Other questions that might arise in the group decision making context are also discussed. For example, even if a single individual serves as the decision maker, he or she will often consider requesting advice from one or more individuals in the form of probabilities and/or utilities. In such a situation the decision maker would want to know if a group is more likely to provide probabilities and/or utilities that are in some sense "better" than those provided by a single individual. Although little is known about individual versus group judgments of probability and utility, research on other types of human judgments suggests that group judgments will generally be "better" than individual judgments.

II.B.1. Group probabilities. The question of how to arrive at group probabilities is in some sense easier to answer than the same question for utilities. To begin with, more disagreement can reasonably be expected in individual utilities than in individual probabilities. Probabilities should be generated by data and expertise. At a philosophical level, differences in probability assessments should be resolved by accepting the judgment of the person with superior expertise or access to better data, even though practically ascertaining the extent of expertise will be difficult. No such resolution seems to exist for utilities, although, as noted later, certain aspects of utility assessment should possibly depend on expertise.

Even at a more practical level, the determination of group probabilities has an advantage in that more objective methods of validation exist. For example, group probabilities determined by different methods can be calibrated against what actually occurs. In addition, probabilities can be assessed in situations for which an appropriate calculation based on a known data-generating process yields the "right" answer. No similar objective validation methods are available for judging the quality of group utilities. Because of the known theoretical difficulties in determining group probabilities and utilities, these measures of quality become very important in deciding by what procedures group probabilities should be determined.

Two general approaches to arriving at group probabilities have been suggested: mathematical approaches in which the probabilities of individuals are combined mathematically into a group probability, and behavioral approaches which depend on interaction and communication among individuals to arrive at a consensus probability or at least reduce disagreement. Of the five mathematical approaches reviewed by Seaver, the simplest and most widely used is the weighted linear combination; that is, averaging. Because of its simplicity, its general acceptability to groups, and the lack of research showing that any of the other approaches are better, this is probably the most feasible of the mathematical approaches.

Although a wide range of behavioral approaches are possible, two have been the subject of considerable research and actual use. Both the Delphi method and the nominal-group method depend on well-structured communications to transmit information among group members. This communication should, and usually does, increase agreement, but will not usually result in a consensus. Therefore, some mathematical combination rule must still be used. The primary difference between the two techniques is that the nominal-group method uses structured face-to-face interaction to transmit information among group members, while the Delphi method relies on written feedback without group members ever directly interacting. The scant research that has been done to date comparing these two methods for eliciting group probabilities tends to favor the nominal-group technique. Obviously, more research is called for before any well-established guidelines concerning how to determine group probabilities will be available.

II.B.2. Group utilities. Several formal procedures for combining individual preference or utility functions into a group function were also reviewed by Seaver. All suffered from some disadvantage such as very restricted applicability or violation of Pareto optimality. Edwards (see summary No. 2) has suggested a more heuristic approach to determining "public" utilities based on a simple multiattribute rating technique (SMART). Two types of judgmental inputs are needed for SMART: ratings of the utility of alternatives on each single attribute and ratio judgments of the importance of each attribute. Single attribute utilities can often be determined by objective information or, if not, they should be determined by the judgments of appropriately selected experts. Therefore, differences in these judgments will usually be primarily differences in expertise. If real experts have been selected for these judgments, such differences should be relatively small and unimportant. When large differences exist, they should be resolved in the manner in which differences in expertise are usually resolved: namely, by using the judgments of the best expert(s).

Real and important differences in utilities will usually be exhibited in the importance weights. SMART allows such differences to be explicitly known and discussed. By focussing on the exact nature of disagreement, strongly professed differences in opinion may be much reduced.

Edwards reviewed three examples of how SMART has been used to determine public values in the evaluation of building permits in the California coastal zone, possible research programs in the Office of Child Development, and water quality for fish and wildlife and human consumption. Probably the most striking example of how SMART can reduce disagreement about values came in the building permit example taken from Gardiner's (1974) PhD dissertation. This example compared the evaluations of two groups. One group consisted of self-reported conservationists, while members of the other group tended to be more development oriented. When holistic, intuitive evaluations of building permits were made, the groups differed substantially. Yet when the evaluations were made via SMART, the differences were greatly reduced. A likely explanation for this reduction in disagreement seems to be that holistic evaluations allow people with strong viewpoints to concentrate on the aspects of the entities being evaluated about which they feel most strongly. On the other hand, SMART does not allow concentration on a few controversial attributes. Agreement on several non-controversial attributes will lessen the overall disagreement caused by differences on the controversial attributes. SMART, of course, will not eliminate all disagreements, but when it does not, it will focus attention on the real points of disagreement.

Some of the work reviewed in the previous section is also relevant to the consideration of public utilities. If, as has been suggested above, real differences of opinion will be expressed in the importance weights assigned to attributes in SMART, our work with equal weights would be relevant to this problem. By showing the conditions under which evaluation using equal weights is a good approximation to evaluation using differential weights, we can provide guidelines for when equal weights can be used for public utilities. This would eliminate the need to resolve differences in opinion about the importance of attributes, since such differences would have little effect on the final evaluations.

II.C. Elicitation and Quantification of Uncertainty

Until the last few years, research on the assessment of uncertainty (probabilities) was far more common than research on the assessment of subjective values (utilities). Therefore, much of the research on subjective probabilities is oriented toward particular problems of application rather than the theoretical problems encountered in utility research.

One of the practical problems with which we must contend is that in many situations, assessed probabilities consistently violate certain rules of probability theory. We are now working toward the development of methods that will reduce or eliminate these consistent errors. Thus, although we are increasingly knowledgable about what the problems are in assessing subjective uncertainty, many of the practical solutions remain to be found.

II.C.1. The effect of response scales on odds judgments. One finding that has been fairly consistent throughout research on the elicitation and quantification of subjective uncertainty is that different methods of asking the same question produce different responses. For example, in opinion revision studies, responses on logarithmic scales of odds are less conservative than odds simply written or stated verbally (Goodman, 1973; Phillips and Edwards, 1966). A pilot study run prior to the experiment reported by Seaver, von Winterfeldt, and Edwards (1975) also suggested that the endpoints of a response scale may affect the responses. Although they were not systematically manipulated, odds scales with endpoints of 1000:1 seemed to elicit odds judgments that were generally lower than responses on scales with endpoints of 10,000:1.

To further explore exactly how response scales affect odds judgments, Stillwell, Seaver, and Edwards (see summary No. 8) systematically investigated two factors: linear versus logarithmic spacing; and endpoints of 100:1, 1000:1, and 10,000:1. Also included was a response mode in which subjects simply filled in a blank with their odds. Responses on logarithmically spaced scales were found to be superior both to responses on linearly spaced scales and to written odds on all measures of the quality of judgments. The question of the effect of the endpoints was less clear. Responses on scales with 1000:1 endpoints were superior on some measures and responses on scales with 10,000:1 endpoints were superior on other measures.

We are currently further exploring the effects of the response scales in a wider variety of contexts by varying d' and the range of the veridical odds. We also hope to reach a more definitive conclusion regarding the effects of the scale endpoints on responses.

II.C.2. Biases in subjective probability distributions. As the use of decision analytic tools has become more sophisticated over the past several years, researchers in the field have recognized that the uncertainty entering into decision problems is often about continuous rather than discrete variables. This has led to research showing the problems associated with the assessment of continuous probability distributions

(or approximations thereof). Much of the research has focussed on the phenomenon of "surprises" in which a high percentage of true values fall into the extreme tails of the assessed distributions. A second bias has also been found in assessed subjective distributions when the unknown variables are percentages: namely, a tendency for distributions to be displaced to the right when the true percentage is low, and displaced to the left when the true percentage is high. The possible existence of another bias has been suggested by the recent experiment of Seaver, et al. (1975). Their results showed an overall tendency for assessed distributions to underestimate the true values.

Fujii, Seaver, and Edwards (see summary No. 7) have explored these biases more thoroughly. In addition, the effects on these biases of using different procedures to elicit the subjective distributions were also investigated. Two elicitation procedures were used: the fractile procedure in which subjects are asked to judge values of the unknown variable that correspond to fixed levels of their cumulative probability distribution, and the odds procedure in which subjects judge cumulative odds for fixed values of the unknown variable.

The results of Fujii, et al. suggest an interaction between the elicitation procedure and the measures used to show the existence of the biases. The underestimation bias was found when the fractile procedure was used but not with the odds procedure. The same is true for a high percentage of surprises. Only the displacement bias for high and low percentages was found for both elicitation procedures used in the study. The displacement bias was also at least partially responsible for the high percentage of surprises. This study substantiates the Seaver, et al. conclusion that use of the fractile elicitation procedure generally produces distributions that are inferior by most measures of quality to those elicited by the odds procedure.

II.C.3. Averaging as a means of probabilistic inference. One of the practical tools that has grown out of the research on assessing subjective probabilities is probabilistic information processing (PIP) systems. PIP systems were developed to alleviate the phenomenon called conservatism. Research showed rather conclusively that people do not revise their probabilistic judgments on the basis of data as much as Bayes' Theorem, the appropriate rule for probabilistic revision, indicated they should. In a PIP system, only judgments of the diagnosticity of single data are made by people; a task research has shown people perform quite well. These judgments are then aggregated via Bayes' Theorem to produce the desired posterior probabilities or odds.

Goodman's (1973) reanalysis of the data from several experiments on PIP suggested a practical difficulty that may arise when using PIP in real-world settings. She found that one of the most important factors contributing to conservatism was feedback of calculated posterior probabilities or odds to subjects making likelihood ratio judgments about single data. In most real-world settings, the posterior probabilities or odds would probably be available to experts making the single datum judgments, therefore, reducing the effectiveness of PIP.

An experiment by Eils, Seaver, and Edwards (see summary No. 6) provides evidence that other processes than PIP may be effective in probabilistic inference tasks. This study compares the usual cumulative posterior odds judgments with a new type of aggregated probabilistic judgment: average certainty for a sequence of data. Use of appropriate instructions and response scales made the average certainty judgments good subjective assessments of the arithmetic mean log likelihood ratio which can then be plugged into the appropriate form of Bayes' Theorem to calculate posterior odds. The results showed that with proper instructions, subjects could make these judgments, and, indeed, made them very well. Posterior odds calculated from the average certainty judgments were very nearly verdical, while the direct posterior odds judgments were, as usual, very conservative. This idea is new and as yet not well-tested. More research is needed to check the effects of varying data diagnosticity and to see how feedback might affect average certainty judgments. However, the striking results of the current research suggest the idea may be worth pursuing.

<u>II.C.4. Assessing very small probabilities</u>. Many societal decisions involve very small possibilities of very large losses. We believe decision analysis is a tool applicable to this type of decision. However, problems arise in the assessment of the probabilities of very unlikely events. Techniques currently used to assess these probabilities depend on the judgments of experts and/or fault tree analysis. Research in other areas, for example the Eils et al. study, has shown that people are biased against expressing extreme judgments to questions about their subjective certainty. This suggests that the judgmental probabilities that have been used in fault trees are probably biased in the same way.

The use of "marker" events with known relative frequencies is an alternative to direct judgments and fault trees. The likelihood of events with unknown probabilities could be compared with the marker events to produce judgments of the form: "This event is more likely than event A but less likely than event B." If the marker events are chosen to provide a fine enough scale, the resulting range of probability should be precise enough for most purposes.

Because of the newness of this concept, our knowledge currently consists primarily of ideas rather than data. Therefore, the following discussion does not represent a _fait accompli_, but rather illustrates our thinking about how to procede in developing a usable set of marker events and testing the feasibility of its use. The only relevant experimental evidence comes from an unpublished study by Slovic, Lichtenstein, and Fischhoff at the Oregon Research Institute. Their findings indicated that if this technique is to be used, care must be taken in the precise implementation methods. Specifically, the marker events must not be ones that evoke biases in their judged relative likelihood.

The process by which such a set of marker events is developed and tested for feasibility will necessarily involve extensive experimentation. First, possible marker events must be identified and tested for consistent biases in their judged relative likelihood. When the set of non-biased events is specified, the marker events will be used to elicit judgments of the likelihood of events with probabilities that are known to the experimenter but unknown to the subjects. This approach will be compared with the use of fault trees and direct judgments. After extensive laboratory testing, the use of marker events will be tried in a real, not yet identified setting. We have high hopes that the marker event approach will prove successful, since we firmly believe fault trees and direct judgments cannot provide an adequate representation of small probabilities.

III. References

Arrow, K. Social Choice and Individual Values. New York: Wiley, 1951.

Collins, B., & Guetzkow, H. A Social Psychology of Group Processes for Decision-making. New York: Wiley, 1964.

Dalkey, N. An impossibility theorem for group probability functions. Rand Paper P-4862. The Rand Corporation, Santa Monica, Ca., 1972.

Edwards, W. Social utilities. The Engineering Economist, 1972, 6, 119-129.

Edwards, W. Research on the technology of inference and decision. Technical Report No. 011313-F, Engineering Psychology Laboratory, University of Michigan, November, 1973.

Edwards, W. Research on the technology of inference and decision. Research Report No. 75-10, Social Science Research Institute, University of Southern California, August, 1975.

Einhorn, H. & Hogarth, R. Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Fischer, G. Four methods for assessing multi-attribute utilities: An experimental validation. Technical Report No. 037230-6-T, Engineering Psychology Laboratory, University of Michigan, September, 1972.

Fishburn, P. Approximations of two-attribute utility functions. Mimeographed manuscript, College of Business Administration, Pennsylvania State University, 1976.

Gardiner, P. Public policy decision making: The application of decision technology and Monte Carlo simulation to multiple objective decisions--A case study in California Coastal Zone management. Unpublished doctoral dissertation, University of Southern California, 1974.

Goodman, B. Direct estimation procedures for eliciting judgments about uncertain events. Technical Report No. 011313-5-T, Engineering Psychology Laboratory, University of Michigan, November, 1973.

Phillips, L. & Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-357.

Seaver, D., von Winterfeldt, D., & Edwards, W. Eliciting subjective probability distributions on continuous variables. Research Report 75-8, Social Science Research Institute, University of Southern California, August, 1976.

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. <u>Psychological Bulletin</u>, 1976, <u>83</u>, 213-217.

IV. Summaries of Technical Reports

Summary No. 1

How to Use Multi-Attribute Utility Measurement for Social Decision Making

Ward Edwards

Decisions do, and should, depend on values and probabilities--both subjective quantities. Public decisions, even more than other kinds, also should depend on values and probabilities. These quantities should be public, not only in the sense of being publishable, but also in the sense that the values, and perhaps the probabilities, that lie behind the decision should depend on some kind of social consensus, or at least on some kind of aggregation of individual views, rather than on any single individual's views.

The thrust of this paper is that a public value is a value assigned to an outcome by a public, usually by means of some public institution that does the evaluating. This amounts to treating "a public" as a sort of organism whose values can be elicited by some appropriate adaptation of the methods already in use to elicit individual values. From this point of view, the interest of the problem lies in finding the appropriate adaptation of those methods, an adaptation that will take into account individual disagreements about values, individual differences in relevant expertise, existing social structures for making public decisions, and problems of feasibility.

Arguments over public policy typically turn out to hinge on disagreements about values. Such disagreements are often about degree, not kind; developed and developing nations may agree on the vitures both of increased industrialization and decreased degradation of the environment, but may differ about the relative importance of these goals. Normally, such disagreements are fought out in the context of specific decisions, over and over again, at enormous social cost each time another decision must be made.

Multi-attribute utility measurement can spell out explicitly what the values of each participant (decision-maker, expert, pressure group, government, etc.) are, show how much they differ, and in the process can frequently reduce the extent of such differences. The exploitation of this technology permits regulatory or administrative agencies and other public decision-making organizations to shift their attention from specific actions to the values these actions serve and the decision-making mechanisms that implement these values. By explicitly negotiating about, agreeing on, and (if appropriate) publicizing a set of values, a decision-making organization can, in effect, inform those affected by its decisions about its ground rules. This can often remove the uncertainty inherent in planning, and can often eliminate the need for costly, time-consuming, case-by-case adversary or negotiating proceedings. Thus, explicit social policies can be defined and implemented with more efficiency and less ambiguity. Moreover, such policies can easily be changed in response to new circumstances or changing value systems, and information about such changes can be easily, efficiently, and explicitly disseminated, greatly easing the task of implementing policy change.

The paper is structured around three examples. One is land use management; the specific example will be a study aimed at the decision problems of the California Coastal Commission. The decision-making body in this case is a regulatory agency exposed to a wide variety of social pressures from those with stakes in its actions.

The second example is concerned with administrative decision-making; specifically, with the process that the Office of Child Development of the U. S. Department of Health, Education, and Welfare used to develop its research program for the 1974 fiscal year.

The third example is more abstract; it concerns an attempt to develop a consensus among disagreeing experts on water quality, about a measure of the merits of various water sources for two purposes: the input, before treatment, to a public water supply, and an environment for fish and wildlife.

The focus of this paper is on planning. I do not understand the differences among evaluations of plans, evaluations of ongoing projects, and evaluations of completed projects; all seem to me to be instances of the same kind of intellectual activity. Multi-attribute utility measurement can and, I believe, should be applied to all three; the only difference is that in ongoing or completed projects there are more opportunities to replace judgmental estimates of locations on value dimensions with utility transforms on actual measurements--still subjective, but with firmer ground in evidence.

Summary No. 2

Assessment of Group Preferences and Group Uncertainty for Decision Making

David A. Seaver

Decision analysis has rapidly become an accepted tool for aiding decision makers to make optimal decisions. The use of decision analysis involves the quantification of the decision maker's preferences and opinions as utilities and subjective probabilities respectively. However, the formal theory underlying the development of decision analysis is based on the decision maker being a single identifiable individual. Often groups rather than individuals serve as decision makers. Even when a single individual functions as the decision maker, a group may be called upon to provide the inputs necessary for making decisions. In these situations, group utilities and probabilities must be determined. The obvious approach to determining group utilities and probabilities is to somehow combine the judgments of the individuals in the group into a group judgment. Theoretical research, however, has proved that no really satisfactory method for combining individual utilities or probabilities into a group utility or probability exists. The purpose of this report is to explore the possibilities that exist for determining group utilities and probabilities, focussing on the advantages and disadvantages of the various procedures.

The report begins by assessing the current state of the art with respect to determining group preferences and utilities. Three specific possible methods for combining individual preference or utility functions into group preference or utility functions are explored. All suffer from rather severe disadvantages such as restrictive applicability or violation of Pareto optimality. Certain experimental conditions that may reduce disagreement and, therefore, lead to a greater chance of unanimity among group members are also discussed.

There are two general procedures for forming group probability judgments: mathematical aggregation procedures and behavioral methods. The mathematical aggregation procedures depend on a mathematical formula for determining the group probabilities from the individual probabilities. Several possibilities exist, but those with the best underlying theory typically cannot be used in practical situations because of the difficulty in determining some of the necessary inputs.

The behavioral methods utilize interaction and/or communication among the group members to try to reduce the disagreement among group members so a consensus will result. The most widely used methods depend on highly structured communication to allow the group to profit from certain advantages of group interaction that are well-documented by social psychological research.

Since none of the procedures reviewed for forming group utilities or probabilities is completely acceptable on a theoretical level, choice among any set of applicable procedures should be based on empirical observations of the quality of the resulting group judgments. However, very little empirical research has been done in this area, so until more research is done, few conclusions about the relative effectiveness of the different methods can be drawn.

Summary No. 3

Unit Versus Differential Weighting Schemes for Decision Making: A Method of Study and Some Preliminary Results

J. Robert Newman, David A. Seaver, and Ward Edwards

A persistent problem in prediction studies and decision making problems is that of weighting the attributes or dimensions of information assumed relevant to the prediction or decision problem. Intuition and past experience has indicated that the attributes should be differentially weighted with the more important ones receiving higher weights. Recently, however, several empirical and theoretical studies has indicated that there are many situations in which differential weighting may not be necessary and that simple unit weighting, i.e., just adding up the attributes of information may be as good as and in some cases better than differential weighting. The implications of this result, if true, have extraordinary practical and theoretical significance, and the problem of weighting requires very careful study.

In this report, the first of a series, a method of generating realistic data is described, and illustrations of how the method can be used to study the usefulness of different data analysis and prediction are given. The method utilizes a computer simulation which generates a N by M data matrix where N is the number of observations and M is the number of variables or measurements taken on each observation. For example N could be 15 automobiles being considered for possible purchase and M could be 10 performance and/or quality factors of importance for each of the automobiles. The entries in the data matrix would be simulated measurement values for each factor on each automobile. The method also allows for the simulation of various types of error in the assigned values. The computer program to accomplish this is outlined. Two examples of the use of the method are given. One compares the familiar multiple regression model with simple unit weighting in a prediction problem to predict a well defined criterion variable from a set of predictor variables. The regression model estimates the weights to be assigned to each predictor whereas the unit weighting model merely adds up the predictors and thus does not assign differential weights. The results indicate that multiple regression is superior to unit weighting for prediction purposes but the differences between the two models are not substantial. The second example compares several ways of forming weighted and unweighted combinations of attributes of dimensions of importance

to help persons make practical decisions. Some of the conditions in which differential weighting is important for practical decision making are specified. The conditions under which differential weighting is not important are also specified.

Summary No. 4

Differential Weighting in Multi-Attribute Utility Measurement: When it Should Not and When it Does Make a Difference

J. Robert Newman

Most important decisions involve choosing among alternatives with multiple value characteristics. A simple ten step procedure has been proposed to help individuals and/or groups make practical decisions. This procedure is called multi-attribute utility analysis. One aspect of the procedure involves assigning importance weights to the attributes or dimensions of importance considered relevant to the decision. Some recent evidence has indicated that such differential weighting may not be necessary and that equal or unit weighting may be as good as far as making the final decision is concerned.

This paper explores some of the conditions under which differential weighting in multi-attribute utility analysis may or may not be appropriate. Two cases are considered: (1) For the case in which the attributes are not related or are related in a positive fashion (non-negatively correlated attributes), and under conditions when no well defined criterion variable is available, differential weighting is not important. Unit or equal weighting will do just as well in the decision analysis. This means, for this case, multi-attribute utility analysis becomes even simpler since the weighting process need not be carried out. However, decision makers may wish to retain a form of weighting during the initial phase of the analysis since this sometimes helps in defining what attributes should be included in the analysis. In other words, differential weighting may have psychological advantages even though nothing is to be gained numerically. (2) For the case of some or all of the attributes being negatively correlated, i.e., more on one attribute means less on some other attribute then differential weighting can make a difference. Thus, the final decision choice can be different when different weighting schemes are used.

An example of case 2 is given for the decision problem of choosing a "best" automobile from a set of automobiles. Some of the attributes considered important for making this decision might be such things as fuel economy, small exterior size, passing/acceleration ability, low interior noise, and so on. These attributes interact and tradeoffs are sometimes necessary. For example, in order to obtain excellent fuel economy, it might be necessary to sacrifice

acceleration. This could be accomplished by considering lighter cars but this, in turn, could adversely affect ride quality, interior size, and so on. It was demonstrated that under these conditions three different weighting schemes led to different automobiles being considered as the "best".

The practical and theoretical implications of this result are discussed. For the case of negatively correlated attributes or a mixture of positive and negative correlations among the attributes differential weighting makes a difference and in practical situations this difference can be very important. This raises the intriguing question of just what weighting scheme should be used since this choice can critically effect the final outcome. Unfortunately, there is no theory to guide our thinking here. Research is continuing into developing such a theoretical rationale and empirical studies such as the one described in the report are also continuing.

Summary No. 5

Experimental Tests of Independence Assumptions For Risky Multiattribute Preferences

Detlof von Winterfeldt

The purpose of this experiment was to analyze models of human preferences in complex decision situations that are characterized by uncertainty and multiple attributes of outcomes. Four basic models for such risky multiattribute preferences were considered, among them the additive and multiplicative expected utility models. Independence assumptions that can test the descriptive validity of these models were formulated.

The validity of the independence assumptions, including the marginality assumption and utility independence, was tested for subjects' preferences among even chance gambles for commodity bundles containing gasoline and ground beef. Subjects matched gambles or commodity bundles against a standard and these matches were checked to see if the indifference held in various stimulus contexts as required by the independence assumptions. Effects of response modes, instructions, and personal preference characteristics were examined.

All independence assumptions and models were violated by a bias to prefer a gamble or commodity bundle that was previously matched against a standard, independently of context. Systematic and strong violations of the marginality assumption were found in the form of a multivariate risk aversion: subjects tended to prefer a gamble with more balanced multiple outcomes over a gamble with extreme multiple outcomes, even if all single outcomes had an equal chance of occurring. Both the bias and multivariate risk aversion were independent of response modes and instructions. Other preference characteristics such as single attribute risk attitude and preferential interaction of commodities seemed unrelated to multivariate risk aversion.

The bias to prefer a previously matched gamble over a standard cannot be explained by any traditional model describing risky multiattribute preferences. This bias could be due either to mismatching or to a change in preferences after matching. The phenomenon of multivariate risk aversion proved to be a stable property of risky multiattribute preferences for the stimuli considered. Descriptive models for risky multiattribute preferences will have to take this phenomenon

into account in similar stimulus situations. For normative modelling, the results of the experiment indicate the necessity to carefully check the consistency of preferences assessed by procedures that are based on indifference judgments and to compare them with actual choices. The multivariate risk aversion effect suggests that simple additive expected utility models may, in some cases, be inappropriate for prescribing preferences. Checks of the marginality assumption and analyses of the form of multivariate risk aversion should be designed and tested carefully, before modelling decision makers' preferences with additive expected utility models.

Summary No. 6.

Developing the Technology of Probabilistic Inference: Aggregating by Averaging Reduces Conservatism

Lee C. Eils, III, David A. Seaver, and Ward Edwards

A relatively large body of research indicates that people are conservative processors of probabilistic information. Recent attention has focussed on two possible explanations of this phenomenon. The misaggregation hypothesis depicts conservatism as an inability to properly combine the information in sequences of data. The other explanation suggests conservatism is the result of a response bias: the avoidance of extreme odds or probability judgments.

This experiment explores the use of a specific response, average certainty, devised to thwart conservatism caused by either a response bias or a certain form of misaggregation. Use of appropriate instructions and response scales made the average certainty judgments good subjective assessments of the arithmetic mean likelihood ratio which can then be plugged into the appropriate form of Bayes' Theorem to calculate posterior odds. These judgments seemed likely not to be affected by a response bias since extreme responses were not needed. In addition, research has suggested that people are more likely to aggregate information by averaging than by adding or multiplying, so misaggregation may be exhibited only in specific forms of aggregation and may not be present in averaging.

Our results indicated that average certainty judgments were both more orderly and more veridical than cumulative certainty judgments of the type usually obtained in probabilistic inference tasks. The cumulative judgments were very conservative while the average certainty judgments were only slightly radical. Since this study was undertaken only to see if average certainty judgments were an effective way to reduce conservatism, it does not directly test what causes conservatism. However, some implications concerning the nature of conservatism are discussed, along with the implications for the technology of probabilistic inference.

Summary No. 7

New and Old Biases in Subjective Probability Distributions: Do They Exist and Are They Affected by Elicitation Procedures?

Tsuneko Fujii, David A. Seaver, and Ward Edwards

Past research indicates that people exhibit biases in assessing probability distributions on continuous variables. Three types of biases have been identified: too many true values falling into the extreme tails of the distributions, a displacement toward 50% for distributions assessed on percentages, and a general tendency to underestimate. This study explored the nature of these biases with particular emphasis on how they interact and how they are affected by the procedure used to elicit the distributions.

Two procedures were used to elicit subjective probability distributions on percentage variables. In the fractile procedure subjects were asked to judge values of the unknown percentage that corresponded to fixed levels of their cumulative probability distribution, while in the odds procedure subjects judged the cumulative odds for fixed values of the unknown percentages. For all the unknown percentages, p%, distributions were assessed for both p% and 1-p%. The extent to which these assessments summed to less than 100% indicated a bias toward underestimation.

Underestimation was generally found when the fractile elicitation was used but not when the odds procedure was used. Also, too many true values fell into the extremes tails of the distributions elicited by the fractile procedure, but no similar bias was found in distributions elicited by the odds procedure. The displacement toward 50% was found in distributions elicited by both procedures. This bias also appeared to be the cause of a considerable number of the true values in the extreme tails of the distributions. Many of the differences in the biases found when different elicitation procedures were used can probably be accounted for by subjects avoiding extreme responses and odds judgments between 1:1 and 2:1.

Summary No. 8

The Effects of Response Scales on Likelihood Ratio Judgments

William G. Stillwell, David A. Seaver, and Ward Edwards

Different methods of recording responses to the same question have been shown to produce different responses. In order to systematically study how response scales affect likelihood ratio judgments, this experiment manipulated two independent variables: the endpoints of the response scales (100:1, 1000:1, 10,000:1) and the spacing of the scales (logarithmic versus linear). Results compared the veridicality of responses on the six scales produced by crossing these factors plus another response mode in which subjects simply wrote their judgment in a blank (no scale).

Logarithmic scales produced responses that were both more veridical and more consistent than responses on linear scales which were, in turn, better than simple written responses. Measures of the effect of the endpoints were somewhat inconsistent and probably interacted with the range of veridical likelihood ratios used in this study. Judgments of relatively small likelihood ratios seemed to be affected by the spacing: linear spacing caused overestimation. Judgments of relatively large likelihood ratios were controlled more by the endpoints: higher endpoints produced larger judgments. Apparently, subjects use the range of the scale as information about the range of true likelihood ratios.

Research Distribution List

Department of Defense

Assistant Director (Environment and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: Lt. Col. Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Office of the Assistant Secretary of Defense (Intelligence)
Attention: CDR Richard Schlaff
The Pentagon, Room 3E279
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Program Management Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209
(two copies)

Administrator, Defense Documentation Center
Attention: DDC-TC
Cameron Station
Alexandria, VA 22314
(12 copies)

Department of the Navy

Office of the Chief of Naval Operations (OP-987)
Attention: Dr. Robert G. Smith
Washington, DC 20350

Director, Engineering Psychology Programs (Code 455)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
(three copies)

Assistant Chief for Technology (Code 200)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)
Attention: Dr. Bruce McDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (ONR)
International Programs (Code 1021P)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye and Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106
(two copies)

Dr. M. Bertin
Office of Naval Research
Scientific Liaison Group
American Embassy – Room A-407
APO San Francisco 96503

Director, Naval Research Laboratory
Technical Information Division (Code 2627)
Washington, DC 20375
(six copies)

Director, Naval Research Laboratory (Code 2029)
Washington, DC 20375
(six copies)

Scientific Advisor
Office of the Deputy Chief of Staff
for Research, Development and Studies
Headquarters, U.S. Marine Corps
Arlington Annex, Columbia Pike
Arlington, VA 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Headquarters, Naval Material Command
(Code 0344)
Attention: Mr. Arnold Rubinstein
Washington, DC 20360

Naval Medical Research and Development
Command (Code 44)
Naval Medical Center
Attention: CDR Paul Nelson
Bethesda, MD 20014

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Naval Personnel Research and Development
Center
Management Support Department (Code 210)
San Diego, CA 92152

Naval Personnel Research and Development
Center (Code 305)
Attention: Dr. Charles Gettys
San Diego, CA 92152

Dr. Fred Muckler
Manned Systems Design, Code 311
Navy Personnel Research and Development
Center
San Diego, CA 92152

Human Factors Department (Code N215)
Naval Training Equipment Center
Orlando, FL 32813

Training Analysis and Evaluation Group
Naval Training Equipment Center
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Department of the Army

Technical Director, U.S. Army Institute for the
Behavioral and Social Sciences
Attention: Dr. J.E. Uhlaner
1300 Wilson Boulevard
Arlington, VA 22209

Director, Individual Training and Performance
Research Laboratory
U.S. Army Institute for the Behavioral and
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Director, Organization and Systems Research
Laboratory
U.S. Army Institute for the Behavioral and
Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Department of the Air Force

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Robert G. Gough, Major, USAF
Associate Professor
Department of Economics, Geography and
Management
USAF Academy, CO 80840

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Aerospace Medical Division (Code RDH)
Attention: Lt. Col. John Courtright
Brooks AFB, TX 78235

Other Institutions

The Johns Hopkins University
Department of Psychology
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Director, Human Factors Wing
Defense and Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsville, Toronto
Ontario, Canada

Stanford University
Attention: Dr. R.A. Howard
Stanford, CA 94305

Montgomery College
Department of Psychology
Attention: Dr. Victor Fields
Rockville, MD 20850

General Research Corporation
Attention: Mr. George Pugh
7655 Old Springhouse Road
McLean, VA 22101

Oceanautics, Incorporated
Attention: Dr. W.S. Vaughan
3308 Dodge Park Road
Landover, MD 20785

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A.D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Catholic University
Attention: Dr. Bruce M. Ross
Washington, DC 20017

Stanford Research Institute
Decision Analysis Group
Attention: Dr. Allan C. Miller III
Menlo Park, CA 94025

Human Factors Research, Incorporated
Santa Barbara Research Park
Attention: Dr. Robert R. Mackie
6780 Cortona Drive
Goleta, CA 93017

University of Washington
Department of Psychology
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Eclectech Associates, Incorporated
Post Office Box 179
Attention: Mr. Alan J. Pesch
North Stonington, CT 06359

Hebrew University
Department of Psychology
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. T. Owen Jacobs
Post Office Box 3122
Ft. Leavenworth, KS 66027

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
001813-5-F

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Research on the Technology of Inference and Decision

5. TYPE OF REPORT & PERIOD COVERED
Final Technical
7/75 - 9/76

6. PERFORMING ORG. REPORT NUMBER
SSRI 76-7

7. AUTHOR(s)
Ward Edwards and David A. Seaver

8. CONTRACT OR GRANT NUMBER(s)
Prime Contract N00014-76-C-0074
Subcontract 75-030-0711

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Social Science Research Institute
University of Southern California
Los Angeles, CA 90007

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

12. REPORT DATE
October, 1976

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
Decisions and Designs, Inc.
Suite 100, 7900 Westpark Drive
McLean, Virginia 22101
(under contract from Office of Naval Research)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

multiattribute utility
validation
simulation
approximation
group probabilities
group utilities
subjective probability
response modes
biases
elicitation procedures
small probabilities
probabilistic inference

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes ~~15 months of~~ research on the technology of inference and decision. The research, ~~reported in eight technical reports,~~ focussed on three major themes: validation of multiattribute utility models and assessment procedures, the assessment of group probabilities and utilities, and the elicitation and quantification of uncertainty. In this report we explain how this research integrates into an overall program ~~of~~ research.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (continued)

were used
~~We have used~~ both experimentation and simulation to find where errors in multiattribute utilities may exist, how extensive they are, and how they affect the final evaluative process. One study showed the additive multiattribute utility model was generally inconsistent with the expressed preferences of subjects. Simulation has shown that certain simple models may be good approximations of more complex models under specific conditions. However, often these conditions may not be present, so care must be taken in using the approximations.

Theoretical results suggest that no entirely satisfactory method exists for combining individual probabilities and utilities into group probabilities and utilities. We have reviewed the advantages and disadvantages of both the mathematical and behavioral methods that have been suggested for forming these group judgments, and propose use of some particular procedures.

We explored the existence of biases in judgments of uncertainty, and how different elicitation procedures may reduce these biases. A particular problem is the quantification of judgment about very unlikely events. We have identified an approach to eliciting these judgments that we feel is superior to methods currently in use. Extensive experimentation will be necessary to develop this method.

Social Science Research Institute
Research Reports

76-1 William E. McGarvey. Can Adjustment Cause Achievement?: A Cross-Lagged Panel Analysis. March, 1976

76-2 Robert M. Carter, Cameron R. Dightman, and Malcolm W. Klein. The System Rate Approach to Description and Evaluation of Criminal Justice Systems. (Reprinted from *Criminology)*.

76-3 Ward Edwards. How to Use Multi-Attribute Utility Measurement for Social Decision-Making. August, 1976.

76-4 David A. Seaver. Assessment of Group Preference and Group Uncertainty for Decision-Making. August, 1976.

76-5 J. Robert Newman, David A. Seaver, and Ward Edwards. Unit Versus Differential Weighting Schemes for Decision Making: A Method of Study and Some Preliminary Results. July, 1976.

76-6 J. Robert Newman. Differential Weighting in Multi-Attribute Utility Measurement: When it should Not and When it Does Make a Difference. August, 1976.

76-7 Ward Edwards and David A. Seaver. Research on the Technology of Inference and Decision. October, 1976.